ORIGINAL PAPER

Orchardgrass (*Dactylis glomerata* L.) EST and SSR marker development, annotation, and transferability

B. Shaun Bushman · Steve R. Larson · Metin Tuna · Mark S. West ·
Alvaro G. Hernandez · Deepika Vullaganti · George Gong · Joseph G. Robins ·
Kevin B. Jensen · Jyothi Thimmapuram

Abstract Orchardgrass, or cocksfoot [*Dactylis glomerata* (L.)], has been naturalized on nearly every continent and is a commonly used species for forage and hay production. All major cultivated varieties of orchardgrass are autotetraploid, and few tools or information are available for functional and comparative genetic analyses and improvement of the species. To improve the genetic resources for orchardgrass, we have developed an EST library and SSR markers from salt, drought, and cold stressed tissues. The ESTs were bi-directionally sequenced from clones and combined into 17,373 unigenes. Unigenes were annotated based on putative orthology to genes from rice, Triticeae grasses, other Poaceae, *Arabidopsis*, and the non-redundant database of the NCBI. Of 1,162 SSR markers developed, approximately 80% showed amplification products across a set of orchardgrass germplasm, and 40% across related *Festuca* and *Lolium* species. When orchardgrass subspecies were genotyped using 33 SSR markers their within-accession similarity values ranged from 0.44 to 0.71, with Mediterranean accessions having a higher similarity. The total number of genotyped bands was greater for tetraploid accessions compared to diploid accessions. Clustering analysis indicated grouping of Mediterranean subspecies and central Asian subspecies, while the *D. glomerata* ssp. *aschersoniana* was closest related to three cultivated varieties.

B. S. Bushman (✉) · S. R. Larson · J. G. Robins · K. B. Jensen
USDA-ARS Forage and Range Research Lab,
695 N 1100 E, Logan, UT 84322-6300, USA
e-mail: shaun.bushman@ars.usda.gov

M. S. West
USDA-ARS, 2150 Centre Avenue, Fort Collins,
CO 80526-8119, USA

A. G. Hernandez · D. Vullaganti · G. Gong · J. Thimmapuram
Roy J. Carver Biotechnology Center, University of Illinois
at Urbana-Champaign, Urbana, IL 61801, USA

M. Tuna
Faculty of Agriculture, Department of Field Crops,
Namik Kemal University, 59100 Tekirdag, Turkey

Introduction

Dactylis L. comprises a monospecific genus in the Festucaea tribe of the Poaceae family, with the main species *D. glomerata* L. referred to as orchardgrass or cocksfoot. It is distinct and isolated from other genera within Poaceae (Catalan et al. 2004), yet has extensive variation for taxonomic characters, and occupies highly diverse habitats (Stebbins and Zohary 1959). Indigenous to Eurasia and northern Africa, *Dactylis* is now naturalized on nearly every continent due to its extensive use for forage and hay, and is one of the top four perennial forage grass genera grown worldwide (Stewart and Ellison 2010). Within the state of Oregon, USA, where the majority of the orchardgrass seed is produced within the USA, approximately 14 million pounds of orchardgrass seed is produced each year at a value of over nine million USD (Oregon Agricultural Information Network).

Because of orchardgrass importance as a forage and hay grass, an understanding of species, population, and cultivar relationships will benefit plant breeding and enhance efforts to improve forage yield, quality, and other value-added traits. Systematic characterization of *Dactylis glomerata* L. has revolved around the number of subspecies and how they

are designated, with a range from 14 to greater than 22 (Domin 1943; Stebbins and Zohary 1959; Borrill 1978; Lumaret 1988). This has been made more complex due to different taxonomic treatments in Europe, Asia, and North America. The most abundant subspecies, including nearly all cultivated varieties, are autotetraploid and considered to fit within three general subspecies designations: *glomerata* (includes nearly all cultivars), *hispanica* [(Roth) Nyman], and *marina* [Borrill] (Borrill 1978). Aside from some isolated and rare reports of hexaploids, the remaining subspecies are diploid or diploid/tetraploid mixtures (Lumaret 1988). The diploids are the progenitors of tetraploid forms through triploid backcrosses and gametic non-reduction (Borrill 1978). Aside from initial collection sites of wild germplasm, little if anything is known about the progenitors of commonly cultivated varieties of orchardgrass.

A critical weakness in assessing orchardgrass variation and distribution is the lack of molecular marker resources. Previous studies have used a variety of tools: chloroplast and ITS sequences have been used to describe phylogenetic relationships (Lumaret et al. 1989; Catalan et al. 2004; Stewart and Ellison 2010), dominant markers (e.g., RAPD, ISSR, SRAP) used in discreet studies (Koelliker et al. 1999; Tuna et al. 2004; Peng et al. 2008; Zeng et al. 2008), and heterologous markers tested from other forage grass species (Saha et al. 2004; Litrico et al. 2009), However, the open-platform dominant markers generally have the inability to add samples to a study without re-analyzing all previous samples due to the large number of bands genotyped per primer. Heterologous marker transferability also decreases as species diverge (Thiel et al. 2003; Zhang et al. 2005; Litrico et al. 2009), and are accompanied by higher likelihood of homoplasy and polymorphism due to mutations that exist among the species (Thiel et al. 2003; Saha et al. 2004; Zhang et al. 2006). Recently, genomic library-derived SSR markers were used to describe relationships among some Asian germplasm (Xie et al. 2010). Genomic SSR markers also tend to have lower transferability and amplification success than EST-derived SSR markers across diverse germplasm (Varshney et al. 2005; Xie et al. 2010). In contrast to these approaches, EST-derived markers developed from within a target species can circumvent these constraints and provide sufficient polymorphism.

High-quality gene sequence libraries within a species, and the development of markers from them, are the foundation for improving the robustness and accuracy of comparative mapping and functional genetics research. As such, EST-based markers are fundamental for the dissection of complex traits (Varshney et al. 2005). As EST-SSRs markers are derived from transcribed regions, they are likely to have a high success of amplification and associated gene annotations (Varshney et al. 2005), and are useful for in silico comparisons to the homologous

sequences in other species (Bushman et al. 2008). EST-SSR markers have been used in genetic and comparative mapping in a number of forage grass species (e.g., Jones et al. 2002b; Warnke et al. 2004; Saha et al. 2005; Alm et al. 2003; Studer et al. 2008). Additionally, EST libraries with long sequence reads constitute crucial reference sequences for further sequencing and functional genomics studies using "Next-Generation" methods.

To improve the genetic resources for orchardgrass and related species, we report herein the development of an EST library and associated SSR markers from orchardgrass. Our objectives were to obtain high-quality EST sequences, compare their homology to other plant sequences, develop SSR markers and assess marker amplification efficiency and transferability, and to predict the SSR marker genetic map positions based on homologous rice (*Oryza sativa* L.) sequences. We further utilize a subset of the SSR markers to compare the relationships among orchardgrass subspecies, and identify subspecies related to three common tetraploid cultivars.

Materials and methods

Tissue preparation for EST libraries

Construction of the orchardgrass EST libraries used plants from three cultivars: 'Latar', 'Paiute', and 'Potomac' (Alderson and Sharp 1994). The library comprised RNA from etiolated seedlings, salt and drought stressed shoots, salt, and drought stressed roots, and cold-treated crowns (Table 1). For the etiolated seedlings, 20–40 seeds of each cultivar were placed in germination pouches, watered with de-ionized water, and germinated in the dark at 25°C for 10 days. Seed coats were removed before freezing in liquid nitrogen, and the same numbers of seedlings from each cultivar were collected into one tube for RNA extraction.

Table 1 The number of unigenes in the orchardgrass EST library, and the numbers of unigenes with significant BLAST hits ($E < 10^{-5}$) to other grass databases

Number of contigs	12,937
Number of singletons	4,436
Number of unigenes	17,373
Number with BLAST hits	
Nr.aa	13,130 (76%)
TAIR	11,269 (65%)
NCBI_Rice	13,084 (75%)
Barley_TC	12,101 (70%)
Wheat_TC	12,910 (74%)
Maize_TC	11,917 (69%)
Sorghum_TC	10,777 (62%)
NCBI_Festuca	9,489 (55%)

For the salt and drought stressed tissues, four seedlings of each cultivar were germinated in a greenhouse in Logan, UT, in 16 oz. plastic cups in a 80:20 sterilized mix of sand:clay, and were subjected to repeated cyclical drought described in Sack and Grumm (2002). All watering of these plants upon initiation of drought treatment used a Peter's 20-20-20 General Purpose fertilizer solution at 300 ppm, appended with 35 mM $CaCl_2$ and 20 mM NaCl to a solution electrical conductivity of 9 dS/m at 25°C. Tissue harvest occurred 10 weeks after emergence and 21 days after cyclical drought treatment initiation, when water levels were approximately 15% soil holding capacity and when leaves had visible loss of turgor. Prior to collection, soil was loosened and roots were washed quickly in tap water before frozen in liquid nitrogen. Shoots (including the crowns) were separated from roots for subsequent RNA extraction and library tagging. The four plants of each cultivar were bulked for RNA extraction, and equal amounts of RNA were later bulked across the three cultivars.

For cold-treated crowns, four plants of each cultivar were planted and maintained under 14 h light, 24:18°C day:night temperatures until five tillers were growing. One plant of each cultivar was maintained under these control conditions and sampled together with the remaining three cold-acclimated plants. The second plant of each cultivar was cold-acclimated at (4°C) for 4 days, under 8 h light and constant temperature. The last two plants of each cultivar were cold-acclimated as above, and then placed at −4°C with 8:16 h light:dark (PAR 12 μmol/m²s) for 48 h in a growth chamber. The control, cold-acclimated, and one of the two frost treated samples were harvested simultaneously at the time of light application in the growth chamber and sunrise in a greenhouse. The final frost treated sample was harvested the same day, but 9 h later or 1 h after the light was application ended. For all crown samples, crowns were separated from shoots and roots with a scalpel, and bulked into one tube for RNA extraction.

Library construction, sequencing, and analysis

Total RNA was extracted from tissue using Trizol reagent (Invitrogen, Carlsbad, CA), and purified through QIAGEN RNeasy Midi columns (QIAGEN, Valencia, CA). Integrity of RNA was validated with denaturing agarose gels. Poly(-A)⁺mRNA was isolated from total RNA using the Oligotex Direct mRNA kit (Qiagen). Double-stranded cDNAs were synthesized using the Superscript™ double-stranded cDNA synthesis kit (Invitrogen). One microgram of poly(A)+ mRNA from each sample was converted to double-stranded cDNA using the Creator Smart cDNA library construction kit (Clonetech, Mountain View, CA). The first-strand cDNAs for the four samples were differentially tagged at the 5′-end with adaptors. Double-stranded DNAs from all

samples were pooled in equal amount, normalized using the Trimmer-direct kit (Evrogen, Moscow, Russia), cloned into pDNR-LIB vector (Clontech), and transformed using DH10B electrocompetent cells (Invitrogen). Upon sequencing 15,264 clones, the cDNA inserts of the sequenced clones were amplified by PCR and subtracted from the normalized cDNA using the Trimmer-direct kit. The subtracted cDNA (i.e., enriched for novel ESTs) was subsequently cloned and transformed as described above.

The normalized and subtracted libraries were sequenced with a custom primer for the 5′-end of the inserts (5′-CG AGCGCAGCGAGTCAGT-3′) and with an anchor-18(T) primer for the 3′ end of the insert (5′-TTTTTTTT TTTTTTTTTTV-3′). Base calling used a Phred quality score cutoff of 20 for a threshold length of at least 200 bp. Vector sequences were detected and masked using the Cross_Match program (http://bozeman.mbt.washington.edu/phrap.docs/phrap.html), and repeat and low-complexity sequences removed using RepeatMasker (Smit et al. 1996–2004). Bacterial, mitochondrial, ribosomal, viral, and other unwanted sequences were filtered out based on BLASTN searches. Resulting filtered, high-quality sequences were assembled into contigs using Paracel Transcript Assemblet (PTA; http://www.paracel.com). The average numbers of ESTs per contig were based on the total number of clean sequences in the forward and reverse directions, from which the singletons were subtracted, and then the resulting number of sequences divided by the number of contigs. Additionally, all contigs were manually inspected for false contigs using the transcriptview program of PTA. The EST information management application (ESTIMA) database was used for curation and dissemination of the ESTs (http://titan.biotec.uiuc.edu) (Kumar et al. 2004).

Annotation for the ESTs was obtained by BLASTX or tBLASTX queries against grass and *Arabidopsis* databases with a cutoff expectation value (E-value) of 10^{-5}. The databases examined were non-redundant and rice protein sequences (BLASTX), and *Festuca* sequences (tBLASTX) from the National Center for Biotechnology Information (NCBI) as of Dec 2007; the Poaceae protein database of Gramene (ftp://ftp.gramene.org/pub/gramene/release26/data/protein/sequence/poaceae_sptr.fa; Sep 2007); the barley, maize, and wheat Gene Indices databases of The Institute for Genomic Research (http://www.tigr.org); and *Arabidopsis* proteins from The *Arabidopsis* Information Resource (TAIR; http://www.arabidopsis.org).

SSR marker development and transferability

Unigene sets of the three libraries were queried for SSRs using SSRFinder (http://www.maizemap.org/bioinformatics/SSRFINDER/; Oct 2009). The SSRFinder parameters were:

amplicon lengths between 80 and 250 bp, primer annealing temperatures of 60°C, and minimum repeat lengths of 12 bp and four repeat motifs.

Amplification success and projected polymorphism of the resulting complete set of orchardgrass SSR primers was tested on plants for three genera. Within *Dactylis*, one plant each from the cultivars Paiute and Potomac, and one plant each from the subspecies *himalayensis* [Domin] (PI 295271) and *aschersoniana* [(Graebn.) Thell.] (PI 372621). A bulk of five plants from four tall fescue [*Festuca arundinacea* (Schreb.) S.J. Darbyshire] sources were used: the cultivar 'Jessup', the cultivar 'Fawn', and two in-house experimental lines that are parents of incipient genetic mapping populations (22-OP and 16-OP). To test the primers within *Lolium*, two plants that are parents to an interspecific *L. perenne* (L.) × *L. multiflorum* (Lam.), MFA × MFB, ryegrass mapping population were used (Warnke et al. 2004). For all samples, DNA was extracted using the QIAGEN 96-well extraction kit and following the manufacturer's recommendations. Polymerase chain reaction used Jumpstart (Sigma-Aldrich, St. Louis, MO) hot-start *Taq* polymerase, following the manufacturer's protocol, except that 1 µM R110-5 dCTP (Perkin Elmer, Waltham, MA) was spiked into each reaction, and amplification products resolved on an ABI3730 (Applied Biosystems, Foster City, CA) at the Utah State University Center for Integrated Biosystems (Logan, UT).

Subspecies diversity analysis

Fifteen accessions of eight orchardgrass subspecies were obtained from three gene banks: the National Plant Germplasm System, USA (NPGS); the Leibniz Institute of Plant Genetics and Crop Plant Research, Germany (IPK); and the Aberystwyth Genetic Resources Information System, UK (ABY). The source country of the accessions represents the location where the seed was collected, which in some cases resulted from historical transplanting outside of the native distribution of the subspecies (e.g., Australia). In addition, certified seed of three cultivars was used as checks for cluster analysis: 'Latar' from Granite Seed (Lehi, UT, USA) lot 30850, 'Paiute' from Granite Seed lot 32201, and 'Potomac' from Wheatland Seed (Clearfield, UT, USA) lot P70-3-PT6.

Nuclear DNA content by flow cytometry was carried out on 10 plants for each accession of the subspecies as per Arumuganathan and Earle (1991). Briefly, 50 mg fresh leaf tissue from young orchardgrass seedlings and 20 mg fresh leaf tissue from *Vicia sativa* (internal standard) were placed on ice in a sterile plastic petri dish. The tissue was chopped into 0.25–1 mm segments in 1 ml solution A [24 ml $MgSO_4$ buffer (ice-cold); 25 mg dithiothreitol; 500 µl propidium iodide stock (5.0 mg propidium iodide in 1.0 ml

double distilled H_2O); 625 µl Triton X-100 stock (1.0 g Triton X-100 in 10 ml double distilled H_2O)]. The solution was filtered through a 35 µm nylon mesh into a microcentrifuge tube and centrifuged at high speed (13,000 rpm) for 20 s. The supernatant was discarded; the pellet was resuspended in 400 µl solution B (7.5 ml solution A; 17.5 µl DNAse-free RNAse) and incubated for 20 min at 37°C before flow cytometry analysis. Samples stained with PI were excited with a 15 mW argon ion laser at 488 nm. Mean DNA content per sample was based on analysis of 1,000 nuclei per sample, using the flow cytometer CYTOMICS FC 500 (Beckman Coulter, Fullerten, CA) at the Central Laboratory of the Medical School of Trakya University, Turkey.

From each of the 18 accessions and cultivars, 14 plants were sampled for DNA extraction. DNA was extracted and PCR conducted as listed above. From these accessions, the within-accession mean genetic diversity values were also estimated from the 33 SSR markers using the similarity index (Dice 1945; Lynch 1990; Leonard et al. 1999). The similarity index is defined as: $S_{xy} = 2N_m/N_x + N_y$, where N_m is the number of marker bands that are shared between individuals x and y, and N_x and N_y are the total number of bands present in each respective individual. Standard errors of the mean similarity values were computed as per Leonard et al. (1999). As the banding patterns of autopolyploid organisms with codominant markers may not express individual's genotypes (Kosman and Leonard 2005), SSR markers were scored as presence/absence binary data. A neighbor-joining dendrogram was also constructed using the average dice similarity values among entries, and bootstrap support for the nodes of the dendrogram was generated through 500 similarity matrix permutations by random sampling of plants with replacement. A consensus tree was constructed using Phylip v3.69 (Felsenstein 2009). Only nodes with bootstrap support greater than 0.70 (70) are denoted. The tree was midpoint rooted using FigTree v1.3.1 (Rambaut Research Group 2009).

Results

Development and annotation of orchardgrass EST library

A normalized EST library was constructed from four orchardgrass tissues: etiolated seedlings, cold-treated crowns, salt and drought treated shoots, and salt and drought treated roots. A total of 65,613 EST clones were sequenced, divided into four tagged tissues. The majority of sequences were obtained from salt and drought stressed root tissue (29,732), while only 1,605 were obtained from

the cold-treated crowns. The etiolated seedling and salt and drought stressed shoot tissues each had 12,000–13,000 clones sequenced. From the EST clones, 17,373 unigenes were formed, comprising 12,937 (74%) contigs and 4,436 (26%) singletons (Table 1). The average number of ESTs in a contig was 4.7, with 73% redundancy. The average length of contigs was 912 bp, and the average length of a singleton was 572 bp. All ESTs are curated and freely available on the University of Illinois Carver Biotechnology Center website (http://titan.biotec.uiuc.edu/), and the NCBI Genbank (accessions HO118416 to HO184029).

The orchardgrass unigenes were compared through BLASTX homology searches to several plant sequence databases, and the majority of unigenes had high homology hits (Table 1). Seventy-six percent of the unigenes had hits to the NCBI non-redundant database, 75% to the NCBI rice database, while approximately 70% of the unigenes had hits to cereal grass databases. Tall fescue, a close relative of orchardgrass, showed the least number of homologous hits with 55%, although this likely an artifact of the incompleteness of the tall fescue database relative to the other databases.

All ESTs were aligned with the physical map position of homologous rice sequences (Table 2), but using a BLASTN search with an E-value cutoff of 10^{-5}. A total number of 12,134 of the 17,373 orchardgrass unigenes had hits to rice genes given these parameters, which was 93% of those 13,084 detected using BLASTX algorithm as shown in Table 1. Rice chromosomes one and three had the highest number of genes with homology to orchardgrass, while the number generally decreased in proportion to the rice chromosome size.

Characterization of SSR markers

The orchardgrass unigenes were used to identify perfect SSRs with a minimum SSR length set to 12 bp and the minimum number of repeats set at 4, and an amplicon length ranging from 80 to 250 nucleotides. Seven percent (1,162) of the unigenes contained SSRs fitting those criteria. These SSR-containing unigenes were distributed among the rice chromosomes similar to the entire unigene set (Table 2). The vast majority of SSRs were tri-nucleotide repeats (Fig. 1), which are enriched in protein-coding regions relative to other SSRs. The next most abundant type of SSR were di-nucleotide repeats, followed by tetra-nucleotide repeats.

Within the *Dactylis* genus, four plants representing cultivars and wild germplasm were selected for SSR primer amplification success testing: one from the cultivar Pauite, one from the cultivar Potomac, one from subspecies *aschersoniana*, and one from subspecies *himalayensis*. Approximately 79% (913) of SSR primers showed clear

Table 2 Orchardgrass EST unigenes and SSRs with rice homology hits ($E < 10^{-5}$), arranged by rice chromosome

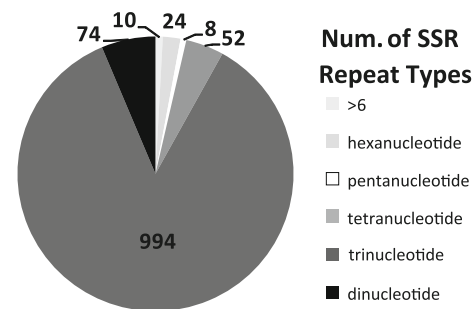| Rice chromosome	*Dactylis* homologous unigenes	*Dactylis* SSRs	Percent
1	1,827	131	7.2
2	1,469	89	6.1
3	1,752	129	7.4
4	1,064	81	7.6
5	1,023	75	7.3
6	949	68	7.2
7	901	68	7.5
8	756	57	7.5
9	620	48	7.7
10	590	41	6.9
11	606	48	7.9
12	577	35	6.1
No rice hit	5,239	292	5.6

Fig. 1 Number of orchardgrass SSRs in each repeat-type

amplification products among the four plants, with *aschersoniana* having 76% success, the two cultivars each having 75% success, and *himalayensis* having 74% success of amplification (Supplemental Table). The remainder of SSR primers showed amplification products larger than the 500 bp capillary gel resolution parameters used, weak amplification, no amplification. When comparing transferability of SSR primer amplification to three related forage grass species, tall fescue (*F. arundinaceum*), perennial ryegrass (*L. perenne*), and annual ryegrass (*L. multiflorum*), 43% (506) of orchardgrass primers amplified strong products among four tall fescue bulk samples, and 39% (457) among the two ryegrass parents of the MFA × MFB interspecific mapping population (Warnke et al. 2004).

Genetic diversity among orchardgrass subspecies with SSR markers

Thirty-three SSR primers with the longest total repeat length were selected for a primer panel to estimate

polymorphism and genetic diversity within the accessions and cultivars (Table 3). Three to 23 bands were detected per primer, with a total of 318 bands (Table 3). The mean number of bands per individual per primer ranged from 4.75 to 0.95 (Table 3), and the average total number of bands per individual (on an accession basis) ranged from 63.4 to 102.4 (Table 4). Of the 33 primer sets, 29 had hits to putative rice orthologs, distributed among 10 of the 12 rice chromosomes. Three cultivars and eight *Dactylis* subspecies were tested for their average within-accession similarity index values (Table 4). Values ranged from 0.44 [*lusitanica* (Stebbins and Zohary (1959) accession) to 0.71 (*juncinella* [(Bory) Boiss.] accession]. The three cultivars had similarity values between 0.59 and 0.64, which was intermediate with respect to the other accessions (Table 4). Two diploid *aschersoniana* accessions had some of the highest similarity values (0.70), indicating less within-accession variation relative to the other accessions.

Using an average similarity based neighbor-joining dendrogram; the three cultivars were closely related with tetraploid *aschersoniana* accessions (Fig. 2; Table 4). Within that clade, Potomac and Paiute showed strong distinct grouping while the node separating Latar from the tetraploid *aschersoniana* accessions did not have bootstrap support. The diploid *aschersoniana* accessions (GR 667 and GR 668) grouped together, and then grouped with the sole *himalayensis* accession, forming a subgroup within the cultivars and tetraploid *aschersoniana* accessions (Fig. 2; Table 4). Both *parthiana* [Parker et Borrill] accessions formed a well supported group, and were closely related to the *woronowii* (Ovcz.) (Stebbins and Zohary 1959) subspecies. Among the three *woronowii* accessions, PI 314081 clustered with the *parthiana* accessions, was distinct from the other two *woronowii* accessions, and contained both diploid and tetraploid plants. The four Mediterranean accessions (spp. *ibizensis*, *lusitanica*, *juncinella*, and *santai*) all grouped together, were diploid, and were differentiated from the other cultivars and accessions with strong bootstrap support. The *juncinella* accession had one of the longest branch lengths, along with the *himalayensis* and GR 667 *aschersoniana* accessions, highlighting a relatively large amount of variation between those and other accessions.

Discussion

The gene sequences and markers reported herein represent a valuable resource for genetic and genomic analyses within *Dactylis* and related genera. The main purpose of these resources will likely be for genetic mapping, functional genetics, and genetic diversity studies. Our results show that the markers will be useful for such purposes. As

the unigene sequences were of lengths over 500 bp, most unigenes and SSR markers were found to have homology to other grass genes, most markers amplified across a diverse set of orchardgrass sources, and a large portion of markers amplified across related genera. Tissues for EST sequencing were collected from three treatments, two of which sampled the plants under abiotic stresses and the third as young seedlings. Approximately half of the resulting sequences came from salt and drought stressed root tissues. To our knowledge, this is the first *Dactylis* EST and molecular marker resource published. As next generation sequencing results are incorporated into forage crops, the long-read high-quality ESTs reported herein, including the large majority with 3'UTR sequence due to bi-directional clone sequencing, will prove a valuable asset for reference sequencing and re-sequencing.

The majority of ESTs also have putative orthologs in other grasses and dicots such as wheat, barley, rice, Arabidopsis, and *Festuca* (Table 2). This suggests that the orchardgrass gene sequence resource will be robust for comparative mapping and candidate gene discovery. In addition, the majority of orchardgrass SSR markers were aligned with homologous rice sequences, and the physical map position of those rice sequences allows for predicted map positions to be in orchardgrass. As a member of the Poaceae clade in the grass phylogeny, orchardgrass has a base chromosome number of seven. Several studies have shown the synteny between rice and the seven-chromosome grass species (Gale and Devos 1998; LaRota and Sorrells 2004; Jones et al. 2002a). Within orchardgrass, synteny to other grasses may allow for targeted mapping and association studies.

Of the grass gene databases compared to the orchardgrass ESTs, the nearest taxonomically related species to orchardgrass is tall fescue, which had the least number of putative orthologs to orchardgrass sequences. This is likely explained by the depth of sequencing in the libraries (for example there were 44,556 tall fescue ESTs vs. 97,042 in the Poaceae database), or perhaps the average length of the tall fescue ESTs. When orchardgrass SSR markers were tested for amplification in tall fescue and *Lolium* species herein, less than 50% had clear and repeatable amplification products. This is consistent with previous efforts to take SSR primers from *Festuca* and *Lolium*, and amplify those using orchardgrass DNA (Litrico et al. 2009). These differences between *Festuca* and *Lolium* and orchardgrass suggest a somewhat unique nature of orchardgrass, and further sequence and marker comparisons among these three genera will likely enlighten the cause for lower relative transferability of markers between *Dactylis*, *Festuca*, and *Lolium*. However, despite the relatively low transferability, approximately 40% of the 1,162 orchardgrass SSR primers still showed clear amplification products in the

Table 3 Orchardgrass SSR markers used for molecular characterization of subspecies, location of their associated rice homolog, total number of bands detected across 18 accessions and cultivars of orchardgrass, and average number of bands per plant

SSR-containing Orchardgrass EST	Repeat type	Repeat length	F primer	R primer	Rice chrom.	Rice map start location	BLASTN E-value	Total num. Bands	Ave. bands/plant	Size range (nt)
Dg_Contig1744	GTTTG	20	GCTGTTGCACAGTTAGTAGTTCGT	CTAACACTGACAGCGTGTTCTTCT	Chrom 3	35,043,546	4.8E-40	23	4.75	198–233
Dg_Contig66	GAA	18	AGGTAAACATTGGAGAGAAAGGCT	CTACTAGTTCCGTCATCTCCTGGT	Chrom 7	509,219	8.5E-90	14	3.08	176–186
BG04063B2F01.r1	TCC	12	CCAATAACACTGGACTCTCTTCCT	TGGCATTTGAAGAGTTACCTATGA	Unknown	N/a	N/a	14	3.20	318–343
BG04056B2F02.r1	CTG	18	ATCTCGATGTGGAACGACTTCT	AGCATCTTCATCAGCAGCAG	Chrom 3	33,074,147	9.0E-35	14	2.09	96–122
Dg_Contig330	AGA	21	AAGCTGAGAAGAAATTAACCGACA	ATGTTGATCACCGTCGTGCT	Chrom 8	20,453,914	4.0E-50	14	2.47	166–181
BG04059A1A07.f1	CGG	12	AAACAAGAAAACAAGAGGAATTGG	GAGTACTTCTCCTTGCACCGTC	Chrom 3	35,979,795	5.7E-54	13	2.77	86–121
Dg_Contig3046	CGG	18	GACGACGAACATCTCTAATTTGGT	CTTCTCCTTGTCGAGGATCTTG	Chrom 7	2,679,480	1.4E-47	13	1.63	202–264
Dg_Contig7660	GCC	15	GACAAGCGGAAATGGATGAG	GTCGTCTCTGATGAAGAATTCAGA	Chrom 3	31,431,705	2.7E-29	13	1.31	310–340
Dg_Contig5978	TGC	18	AAAGATCGAGGATGACAGTTCTCT	AAATTAACAACATTGAGTGGAGGC	Chrom 7	28,443,714	1.9E-82	12	1.91	212–229
Dg_Contig3377	CGC	12	GAGAAGAAGGAGCACAAGTACCC	CTCTCACCGCAGTGGATGTT	Chrom 4	24,114,828	4.0E-64	11	3.33	196–207
Dg_Contig4556	TGT	15	TCTGGAGAAATATGCTGAGAGTG	TTGACTTCCTTTCATAGCGTTACA	Chrom 1	4,739,159	4.4E-54	11	1.51	181–214
Dg_Contig10135	CTG	18	ATGAGGAGGAGATAGAGAAGCTCA	ATCTGATGTTATTCCAAGGAAACG	Chrom 2	1,009,943	1.2E-28	11	1.69	203–224
Dg_Contig660	CAA	15	ACGACGGTAGAGGTAAGAATTGAC	CTCTGTGCTCTCCTCATATTGTGT	Chrom 10	22,568,435	1.7E-11	11	2.69	97–110
BG04030A2C10.f1	CGC	15	AGTCAAACAAGCCTCGTCTC	GTCTTCACTCACTCCATTGCTAGA	Unknown	N/a	N/a	11	2.33	154–187
Dg_Contig10764	GCCT	16	AACGTCCGACAGGAGTGTTAAAT	TAGGAGTAGATTGTGCAGGATGAA	Chrom 1	7,802,180	1.9E-38	10	1.53	90–99
Dg_Contig4296	CGG	15	GCAAGACTACGACTCTCACGG	GTAGATGTTGTGCACGATGGAG	Chrom 9	3,772,265	1.6E-16	10	1.50	184–203
Dg_Contig5054	AAGGAG	24	GTAGGGGAAATACAGGAGAAGGTC	AGCACAGCAACTTCATCAGATAAC	Unknown	N/a	N/a	10	1.82	175–205
Dg_Contig11285	TCTT	16	GCAGTACCATTTCAGGTGAG	AATTTCAGCAGAAACGTCAGAAA	Chrom 9	1,697,9315	7.6E-54	10	1.06	118–131
Dg_Contig4110	GAC	18	TCTGAAGACCAATGATACCAAGAA	CTTCTACTCCATCCTCATCATCCT	Chrom 4	32,418,464	1.9E-18	9	1.65	179–193
Dg_Contig6373	GAG	21	ATCGAGATCAGAAGGTCAAAGAAG	GGGTAGAAGCTGAAGGACCAG	Chrom 12	10,419,382	3.7E-47	9	0.95	171–190
BG04046A2B07.f1	CCT	21	CTGCACCTCAAGCAAAGTGG	AGAACAGGAGGAGCAGGAGC	Chrom 2	5,613,792	3.0E-06	9	3.46	179–237
Dg_Contig3546	AAG	18	AGTGTAGAAGCCTAGCTGTTTGCT	AGTTCAGCTGCTGGGAGGACT	Chrom 7	26,653,011	5.1E-45	9	1.92	162–176
Dg_Contig12453	AGAG	16	CTGAGATTCAAAGTCAACTGTCCA	AAGTTACCGGACCCGATCTC	Chrom 4	35,379,822	8.7E-135	7	2.09	91–108
Dg_Contig10487	CGG	18	CACAACCCCGATTACTGCAC	AGAAGTCGTGAGGTTGGAGT	Chrom 4	35,461,544	6.4E-36	6	1.69	148–171
Dg_Contig4921	ATC	15	AAATTTGAGAAAAGAAACGACCAG	ACGCATAGACATAACCGATGTAGA	Chrom 3	11,666,463	1.7E-34	6	2.38	240–260
Dg_Contig4563	AGA	18	GTATCATTCACCATGGAGGACAG	GAATCTCATCGTAGTCGCTGTATG	Chrom 3	31,507,309	6.0E-160	6	2.19	112–132
Dg_Contig10236	CGAG	16	CGGGTTTTAATCCGGTTCTC	GACCTCTGCATCGCCTTGTA	Chrom 4	24,817,299	4.5E-24	5	1.61	91–99
BG04035A2D08.r1	TTTA	16	TGAGCCTACAATCATGTCAGTCTT	ATAGAAGCCCCATAGTGAGAAGTG	Unknown	N/a	N/a	5	1.42	117–128
Dg_Contig3264	ATG	15	TTCGCTGTATCAAGTCTGAAGAAC	GCATCAAGAAACATTTACAGTTGG	Chrom 7	22,003,311	1.3E-49	5	1.79	127–140
Dg_Contig11508	AGTC	16	AATTGAGAGAAAAAGGTAGGTGTGG	CACCACCACTAGTGTACTCACTCC	Chrom 1	42,899,656	3.1E-35	5	3.15	171–181
Dg_Contig12217	GCCGCC	24	GGTCTGCACAAACCACCTCC	CCATTCCTCACTTCGCTTCCT	Chrom 1	24,780,840	9.8E-42	5	1.26	99–114
Dg_Contig4478	CTG	15	TGATTATATTCAATCGGCTACGTG	AAGATCAGTTCCGCAAATTAGAAC	Chrom 11	19,880,636	7.1E-08	4	1.04	208–224
Dg_Contig667	GCA	15	GAAGTAGCCAGCGATGATGAG	TATCACCTAGGCTGGATGCC	Chrom 8	10,960,154	6.0E-48	3	1.56	178–187

Table 4 Orchardgrass cultivars and accessions tested with 33 SSR markers, and their within-accession genetic diversity generated with a similarity index

D. glomerata subspecies	Source/accession	Source country	Ploidy level	Ave. num. bands per individual[b]	Similarity	SE
Glomerata	Latar cultivar	Russia	4x	102.4	0.590	0.020
glomerata	Paiute cultivar	Turkey	4x	87.9	0.636	0.017
glomerata	Potomac cultivar	USA	4x	88.6	0.629	0.016
aschersoniana	IPK GR 667	Poland	2x	70.9	0.704	0.011
aschersoniana	IPK GR 668	Poland	2x	78.5	0.701	0.032
aschersoniana	IPK GR 5930	Sweden	4x	91.5	0.596	0.015
aschersoniana	NPGS PI 372621	Germany	4x	95.9	0.596	0.049
aschersoniana	NPGS PI 420745	Australia	4x	96.1	0.686	0.019
himalayensis	NPGS PI 295271	India	2x/4x[a]	75.2	0.667	0.043
parthiana	ABY-Bc 5368	Iran	2x	73.4	0.696	0.014
parthiana	IPK GR 10666	United Kingdom	2x	75.4	0.646	0.008
woronowii	NPGS PI 538922	Russia	4x	84.5	0.629	0.012
woronowii	NPGS PI 237610	Iran	4x/2x	96.4	0.644	0.017
woronowii	NPGS PI 314081	Former Soviet Union	4x/2x	66.7	0.658	0.065
ibizensis	NPGS PI 237609	Spain	2x	84.1	0.626	0.021
juncinella	NPGS PI 237601	Spain	2x	70.7	0.709	0.018
lusitanica	NPGS PI 237602	Portugal	2x	63.4	0.444	0.087
santai	NPGS PI 368880	Algeria	2x	77.1	0.562	0.035

ABY Aberystwyth genetic resources information system UK.; *IPK* Leibniz Institute of Plant Genetics and Crop Plant Research, Germany; *NPGS* National Plant Germplasm System, USA

[a] For accessions containing both diploid and tetraploid plants, the majority ploidy level is listed first

[b] The total number of bands across all accessions was 318

Festuca and *Lolium* species. These conserved markers provide several hundred primers for comparative genomics, development of conserved orthologous (COS) markers as has been done in Triticeae (Quraishi et al. 2009), and functional genetics among these related forage-grass species.

Approximately 7% of unigenes contained SSR markers given the parameters we designed (Table 2). This number is within the range, or slightly higher, than the average 3.2% seen in some grass species (Kantety et al. 2002) yet lower than genomic enriched libraries or those in other grass species (Bushman et al. 2008). Similar to tall fescue ESTs (Saha et al. 2004), orchardgrass trinucleotide repeats containing C and G in the first two positions were more abundant than those containing A and T, accounting for 33% of all SSRs. However, no single trinucleotide repeat-type accounted for more than 5% of the SSRs (Supplemental Table 1), which is similar to that found in rice and barley (Temnykh et al. 2000; Thiel et al. 2003). Among the 94 dinucleotide repeats, the most abundant motifs, accounting for 25%, were AG/TC.

Subspecies and cultivars of orchardgrass are all able to hybridize and produce viable progeny with varying efficiency (Stebbins and Zohary 1959). Thus, subspecies can

act as a secondary gene pool (Stuczynski 1992). Relationships among the subspecies have been evaluated using morphological traits (Stebbins and Zohary 1959), chloroplast and ITS sequences (Lumaret et al. 1989; Stewart and Ellison 2010), isozymes (reviewed in Lumaret 1988), and flavonoid compound diversity (Fiasson et al. 1987), and each method has shown commonalities and differences. The SSR markers described herein have great resolution within the genus, which can aid in understanding the more recent genetic history of this species. The *aschersoniana* and *himalayensis* Eurasian subspecies were considered some of the oldest in the genus (Stebbins and Zohary 1959), yet in this study, the two subspecies grouped together with high bootstrap values. The *woronowii* and *parthiana* subspecies are partially sympatric with a hybridization zone between, with the latter occupying higher elevations (Borrill and Carroll 1969). These showed a close association using SSR markers (Fig. 2). Among the Mediterranean and subtropical subspecies, the *lusitanica* subspecies grouped with the three other Mediterranean subspecies. Indeed, *lusitanica* grouped with *juncinella* and *santai* (Stebbins and Zohary 1959) accession using SSR markers (Fig. 2), similar to previous morphological characterizations (Stebbins and Zohary 1959), yet *santai* was

Fig. 2 Neighbor-joining dendrogram of orchardgrass subspecies based on an average pairwise similarity distance matrix, with bootstrap support from 500 iterations

0.05 difference (Dice similarity coefficient)
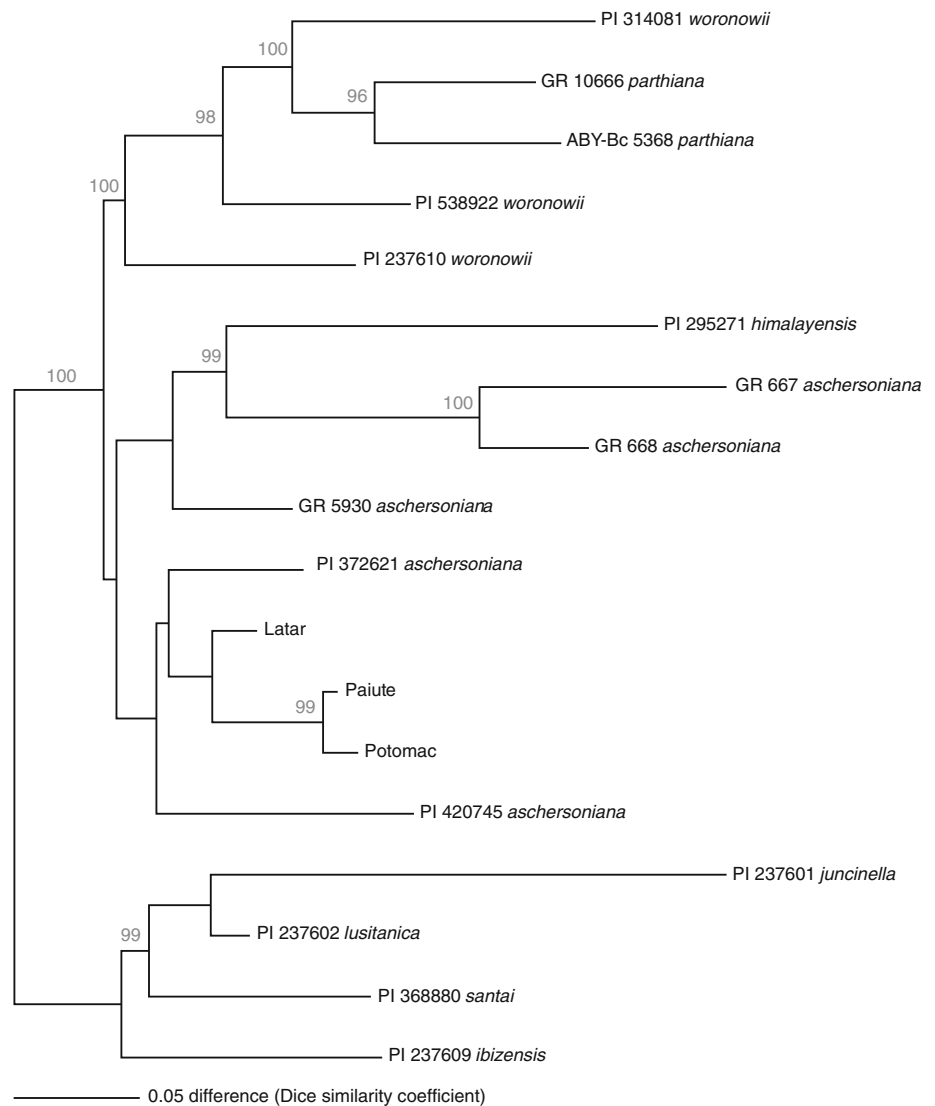
indistinguishable from *aschersoniana* with chloroplast and ITS sequences (Stewart and Ellison 2010). By larger sampling of populations in the diploid subspecies, it will be of further interest to investigate their relationships using these molecular markers.

The *Dactlyis glomerata* ssp. *glomerata* is the most abundant form and includes the common cultivated varieties of orchardgrass. Morphologically, the diploid ancestor of subspecies *glomerata* was considered to be an auto-polyploidization event of *aschersoniana* or of a hybrid between *aschersoniana* and *woronowii* (Myers 1948; Lumaret 1988). In this study, all three orchardgrass cultivars were found to cluster together and group most closely to the tetraploid *aschersoniana* accessions, followed by the tetraploid *woronowii* accessions. These data confirm the implication of *aschersoniana*'s primary role in the ancestry of common cultivated varieties, and are not inconsistent with the possibility of *woronowii* introgression. Although it

may prove difficult to conclusively discern the presence of ancestral *woronowii* due to the ongoing reticulation among diploid and tetraploid populations of orchardgrass (Stebbins and Zohary 1959), these data suggest that *woronowii* would likely provide another suitable source of germplasm for value-added trait introgression into cultivated orchardgrass varieties.

References

Alderson J, Sharp WC (1994) Grass varieties in the United States. USDA-SCS Agric. Handbook 170. US Government Print Office, Washington, DC

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA (2003) A linkage map of meadow fescue (*Festuca pratensis* Huds.) and comparative mapping with other Poaceae species. Theor Appl Genet 108:25–40

Arumuganathan K, Earle ED (1991) Estimation of nuclear DNA content of plants by flow cytometry. Plant Mol Biol Rep 9:229–241

Borrill M (1978) Evolution and genetic resources in cocksfoot. In: Report Welsh Plant Breeding Station, Aberystwyth, UK, 190-209

Borrill M, Carroll CP (1969) A chromosome atlas of the genus *Dactylis* (Part Two). Cytologia 34:6–17

Bushman BS, Larson SR, Mott IW, Cliften PF, Wang RC, Chatterton NJ, Hernandez AG, Ali S, Kim RW, Thimmapuram J, Gong G, Liu L, Mikel MA (2008) Development and annotation of perennial Triticeae ESTs and SSR markers. Genome 51:779–788

Catalan P, Torrecilla P, Rodriguez JAL, Olmsted RG (2004) Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and *trn*L-F sequences. Mol Phlogenet Evol 31:517–541

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302

Domin K (1943) Monograficka studie o rodu *Dactylis* L. Acta bot bohem 14:3–147

Felsenstein J (2009) PHYLIP (Phylogeny Inference Package) version 3.69. Distributed by Author. Department of Genome Sciences: University of Washington, Seattle, WA, USA

Fiasson J-L, Ardouin P, Jay M (1987) A phylogenetic groundplan of the specific complex *Dactlyis glomerata*. Biochem Syst Ecol 15:225–229

Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphres MO, King IP, Kishida T, Yamada T, Balfourier F, Charmet C, Forster JW (2002a) An enhanced molecular marker-based map of perennial ryegrass (*Lolium perenne* L.) reveals comparative relationships with other Poaceae species. Genome 45:282–295

Jones ES, Dupal MP, Dumsday JL, Hughes LJ, Forster JW (2002b) An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 105:577–584

Kantety RV, LaRota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Koelliker R, Stadelmann FJ, Reidy B, Nosberger J (1999) Genetic variability of forage grass cultivars: a comparison of *Festuca pratensis* Huds., *Lolium perenne* L., and *Dactylis glomerata* L. Euphytica 106:261–270

Kosman E, Leonard KJ (2005) Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. Mol Ecol 14:415–424

Kumar CG, LeDuc R, Gong G, Roinishivili L, Lewin HA, Liu L (2004) ESTIMA, a tool for EST management in a multi-project environment. BMC Bioinformatics 5

LaRota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. Funct Integr Genomics 4:34–46

Leonard AC, Franson SE, Hertzberg VS, Smith MK, Toth GP (1999) Hypothesis testing with the similarity index. Mol Ecol 8:2105–2114

Litrico I, Bech N, Flajoulot S, Cadier D, Talon C, Gibelin C, Barre P (2009) Cross-species amplification tests and diversity analysis using 56 PCR markers in *Dactylis glomerata* and *Lolium perenne*. Mol Ecol Rep 9:159–164

Lumaret R (1988) Cytology, genetics and evolution in the genus *Dactylis*. CRC Critical Rev Plant Sci 7:55–91

Lumaret R, Bowman CM, Dyer TA (1989) Autopolyploidy in *Dactylis glomerata* L.: further evidence from studies of chloroplast DNA variation. Theor Appl Genet 78:393–399

Lynch M (1990) The similarity index and DNA fingerprinting. Mol Biol Evol 7:478–484

Myers WM (1948) Studies on the origin of *Dactylis glomerata* L. (Abstract). Genetics 33:117

Oregon Agricultural Information Network (OAIN), Extension Economic Information Office, Oregon State University. Electronic access to this publication is available at: http://oain.oregonstate.edu/

Peng Y, Zhang X-Q, Deng Y, Ma X (2008) Evaluation of genetic diversity in wild orchardgrass (*Dactylis glomerata* L.) based on AFLP markers. Hereditas 145:174–181

Quraishi UM, Abrouk M, Bolot S, Pont C, Throude M, Guilhot N, Confolent C, Bortolini F, Praud S, Murigneux A, Charmet G, Salse J (2009) Genomics in cereals: from genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. Funct Integr Genomics 9:473–484

Rambaut Research Group (2009) FigTree version 1.3.1. Available from tree.bio.ed.ac.uk/software/figtree [accessed April 2010]

Sack L, Grumm PJ (2002) The combined impacts of deep shade and drought on the growth and biomass allocation of shade-tolerant woody seedlings. Oecologia 131:175–185

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Saha MC, Mian R, Zwonitzer JC, Chekhovskiy K, Hopkins AA (2005) An SSR- and AFLP-based genetic linkage map of tall fescue (*Festuca arundinaceae* Schreb.). Theor Appl Genet 110:323–336

Smit ADA, Hubley R, Green P (1996-2004) RepeatMasker Open-3.0

Stebbins GL, Zohary D (1959) Cytogenetics and evolutionary studies in the genus *Dactylis* L. Morphology, distribution and interrelationships of the diploid subspecies. Univ California Publ Bot 31:1–40

Stewart AV, Ellison NW (2010) The genus *Dactylis*; wealth of wild Species: role in plant genome elucidation and improvement, Vol.2. Springer, New York

Stuczynski M (1992) Estimation of the suitability of the Ascherson's cocksfoot (*Dactylis aschersoniana* Graebn.) for field cultivation. Plant breeding acclimatization and seed production (Poland) 36:7–42

Studer B, Asp T, Frei U, Hentrup S, Meally H, Guillard A, Barth S, Muylle H, Roldan-Ruiz I, Barre P, Koning-Boucoiran C, Uenk-Stunnenberg G, Dolstra O, Skot L, Skot KP, Turner LB, Humphreys MO, Kolliker R, Roulund N, Nielsen KK, Lubberstedt T (2008) Expressed sequence tag-derived microsatellite markers of perennial ryegrass (*Lolium perenne* L.). Mol Breeding 21:533–548

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Tuna M, Khadka DK, Shrestha MK (2004) Characterization of natural orchardgrass (*Dactylis glomerata* L.) populations of the Thrace region of Turkey based on ploidy and DNA polymorphisms. Euphytica 135:39–46

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Warnke SE, Barker RE, Jung G, Sim SC, Mian MAR, Saha MC, Brilman LA, Dupal MP, Forster JW (2004) Genetic linkage mapping of an annual x perennial ryegrass population. Theor Appl Genet 109:294–304

Xie W-G, Zhang X-Q, Ma X, Cai H-W, Huang L-K, Peng Y, Zeng B (2010) Diversity comparison and phylogenetic relationships of cocksfoot (*Dactlyis glomerata* L.) germplasm as revealed by SSR markers. Can J Plant Sci 90:13–21

Zeng B, Zhang X-Q, Lan Y, Yang W-Y (2008) Evaluation of genetic diversity and relationships in orchardgrass (*Dactylis glomerata* L.) germplasm based on SRAP markers. Can J Plant Sci 88:53–60

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677–687

Zhang LY, Ravel C, Bernard M, Balfourier F, Leroy P, Feuillet C, Sourdille P (2006) Transferable bread wheat EST-SSRs can be useful for phylogenetic studies among the Triticeae species. Theor Appl Genet 113:407–418